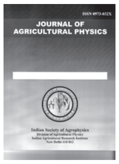
Vol. 23, No. 2, pp. 247-254 (2023)
Journal of Agricultural Physics
ISSN 0973-032X
http://www.agrophysics.in



Research Article

Mustard Yield Prediction at Different Growth Stage using Principal Component Analysis

ANANTA VASHISTH*, ARAVIND K.S., MANOJ KUMAR BECK, MONIKA KUNDU AND P. KIRSHANAN

Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute, New Delhi-110012

ABSTRACT

Mustard yield prediction was done at different growth stage of the crop. Daily weather data during crop growing period as well as mustard yield data for the period of 1984-2019 for ICAR-IARI, New Delhi were used for developing model. Stepwise multiple linear regression and principal component analysis was used for the development of suitable statistical models for multistage mustard yield prediction. Analysis was carried out by fixing 70% of the data for calibration and remaining dataset for validation. Prediction of mustard yield was done at vegetative, flowering and grain filling stage during *Rabi* 2018-19 and 2019-20. The results have revealed that the proposed model can provide reliable pre- harvest prediction of mustard yield in all three stages. Percentage deviation of predicted yield done at vegetative, flowering and grain filling stage by observed yield was 6.35, 9.51 and 6.33% during *Rabi* 2018-19 and 0.17, 6.31 and 7.87% during *Rabi* 2019-20 respectively. On the basis of percentage deviation and model accuracy principal component analysis can be used for predicting mustard yield at different growth stage.

Key words: Weather variables, Stepwise multiple linear regression, Principal component analysis, Yield prediction

Introduction

Mustard is one of the most important oilseed crop grown in *Rabi* season in north-west part of India. Weather variables affect the crop during different stages of development. Thus, the extent of the weather influence on crop yield depends not only on the magnitude of weather variables but also on weather distribution pattern over the full crop season. Hence, predicting crop yield using weather variables is foremost important. Accurate information about history of weather variables and crop yield is useful to take decisions related to agricultural risk management and future predictions. An accurate and timely forecast of crop production with a longer lead time can be very useful, depending on the scale of applications. It allows an agricultural producer to take more informed in-season corrective crop management and financial decision. The government policy maker can take actions such as stocking food supply and strategic resource mobilization in the most insecure areas. Many agricultural industries are increasingly relying on crop market outlooks and yield forecasts for their decision-making. Therefore, there is a need to develop area specific forecast models based on time series data of crop yield and weather parameters with the help of principal component analysis to predict crop yield more accurately. Multiple linear regression has the biggest disadvantage of over-fitting when the number of samples is less than the number of variables. Also, another disadvantage is the multi-collinearity when

*Corresponding author,
Email: ananta.iari@gmail.com

independent predictors are correlated (Verma *et al*., 2016). Garde *et al*. (2015) concluded that stepwise multiple linear techniques can be used effectively for the pre-harvest wheat crop, which are more consistent in performance. Percentage deviation of estimated yield by observed yield by weather based statistical model for maize crop done at flowering stage and at grain filling stage was 10.3 and 7.1% (Vashisth *et al*., 2018). Vashisth *et al*. (2014) reported that percentage deviation of wheat yield prediction done at forty-five days and twenty-five days before harvest by observed yield was less than 10%. Azfar (2015) showed the effectiveness of Principal Component Analysis (PCA) considering all weather indices including interaction indices as regressors was best reliable forecast model for mustard and rapeseed compared to other models. Vashisth and Aravind (2020) reported that Elastic Net, Least Absolute Shrinkage and Selection Operator (LASSO) and Stepwise Multiple Linear Regression (SMLR) model based on weather parameters can be used for multistage mustard yield estimation and Elastic Net performed best among all the three model followed by LASSO and SMLR model. Aravind *et al*. (2022) reported that elastic Net and LASSO was found to be the best model followed by PCA-SMLR, SMLR, Artificial Neural Network (ANN) and PCA-ANN respectively for wheat yield prediction of different location of north-west India. Weather has a great impact on crop yield. The relationship between weather variables and yield of the crop can be estimated though different statistical methods. Mustard yield prediction at different growth stage based on weather variables can be done more accurately by model developed, calibrated and validated with the historical data using principal component analysis. The aim of this study is to evaluate the performance of the models developed for mustard yield prediction at different growth stage using SMLR and PCA-SMLR techniques in order to enhance accuracy of mustard yield prediction.

Materials and Methods

Daily weather data (maximum and minimum temperature, morning and evening relative humidity, rainfall) during mustard crop growing period and mustard yield data for last 35 years were collected from ICAR-IARI, New Delhi. Weather data were arranged in different standard meteorological weeks for three different stages viz. vegetative (40th to 52nd SMW), flowering (40th to 4th SMW) and grain filling (40th to 8th SMW) stage separately. Simple and weighted composite weather indices were developed from the combination of weather variables (Vashisth and Aravind, 2020). 70% of data were used for model calibration and remaining 30% were used for the validation of models. Model for estimating the mustard yield at different growth stage was developed by SMLR and PCA-SMLR techniques.

Stepwise multiple linear regression (SMLR)

Weather indices developed by combination of different weather variables were used for developing model. For developing mustard yield prediction model at different growth stage, yield was taken as dependent variables and simple and weight weather indices along with time was taken as independent variables in SPSS (version 13.0) by SMLR techniques.

Principal component analysis- Stepwise multiple linear regression (SMLR-PCA)

It is a combination of feature selection and selection method used for the data analysis. Principle components scores or factors are calculated from the data analysis which is used as an input variable for stepwise multiple linear regression. It is mainly used to reduce the multicollinearity problem arises from the weather variables. Principal component analysis (PCA) primarily deals with explaining variance through linear combination of original variables.

PCA is very effectively used as a multivariate technique for the purpose of data reduction. PCA transforms the original set of correlated variables in to a new set of uncorrelated variables. If the original variables taken into consideration are uncorrelated, then there so significant meaning in PCA analysis. The first principal component can able to explain the major variability in the data in a greater extend, and remaining each succeeding component accounts rest of the variability as possible.

Let xij be the value of jth weather variable/ biophysical parameter (j=1, 2, ..., P) corresponding to ith treatment of experiment (i= 1, 2, ..., n). The principal component analysis for xij's will be carried out. Let PC1, PC2, ...PCK are the principle component obtained from the analysis. First few PC (K< P) principal component will explain variability (maximum variability) about 90 per cent of the total variation in xij's. Using these K principal components as regressor variables and variety yield (yi) as regress and, the following linear multiple regression model for pre-harvest prediction of crop yield has been proposed.

yi = β0 + β1PC1i + β2PC2i+.......... βkPCki+ei (i= 1, 2, ...n)

where yi is the crop yield of ith variety; β0, β1, β2, .., βk are model parameters and ei denote the error term which is assumed to be follow normal distribution with mean zero and variance σ2. This technique reduces the number of regressors to be used in the model and hence reasonably precise prediction of mustard yield can be obtained even for small set of observation. Flow diagram showing methodology for yield prediction of mustard at different growth stage is shown in Fig. 1.

Model performance during calibration and validation was observed on the basis of root mean square error (RMSE), normalized mean square error (nRMSE) and percent Deviation.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi - Oi)^2}$$

$$nRMSE = \frac{100}{M} * \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi - Oi)^2}$$

$$\text{Percent Deviation (\%)} = \frac{Pi - Oi}{Oi} * 100$$

Where Pi is the predicted value, O_i is the observed value, N is the number of observations, M is the mean of observed value.The prediction is considered excellent with the nRMSE <10%, good if 10–20%, fair if 20–30%, poor if >30%.

Results and Discussion

Mustard Yield prediction by SMLR and PCA-SMLR model at vegetative stage

Model for Mustard yield prediction at vegetative stage for ICAR-IARI, New Delhi was developed by SMLR and PCA-SMLR using long term crop yield data as well as long period daily weather data during sowing to vegetative stage (from 40th to 52th standard meteorological week).

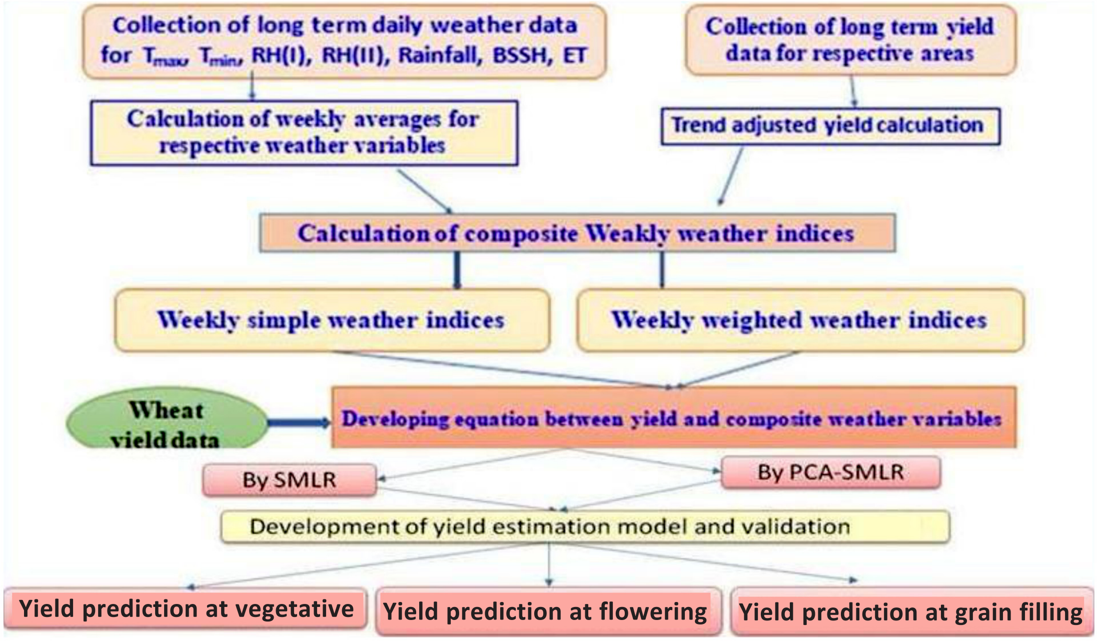

Collection of long term daily weather data for T_{max}, T_{min}, RH(I), RH(II), Rainfall, BSSH, ET
Collection of long term yield data for respective areas
Calculation of weekly averages for respective weather variables
Trend adjusted yield calculation
Calculation of composite Weakly weather indices
Weekly simple weather indices
Weekly weighted weather indices
Wheat yield data
Developing equation between yield and composite weather variables
By SMLR
By PCA-SMLR
Development of yield estimation model and validation
Yield prediction at vegetative
Yield prediction at flowering
Yield prediction at grain filling

Fig. 1. Flow diagram showing methodology for yield prediction of mustard at different growth stage

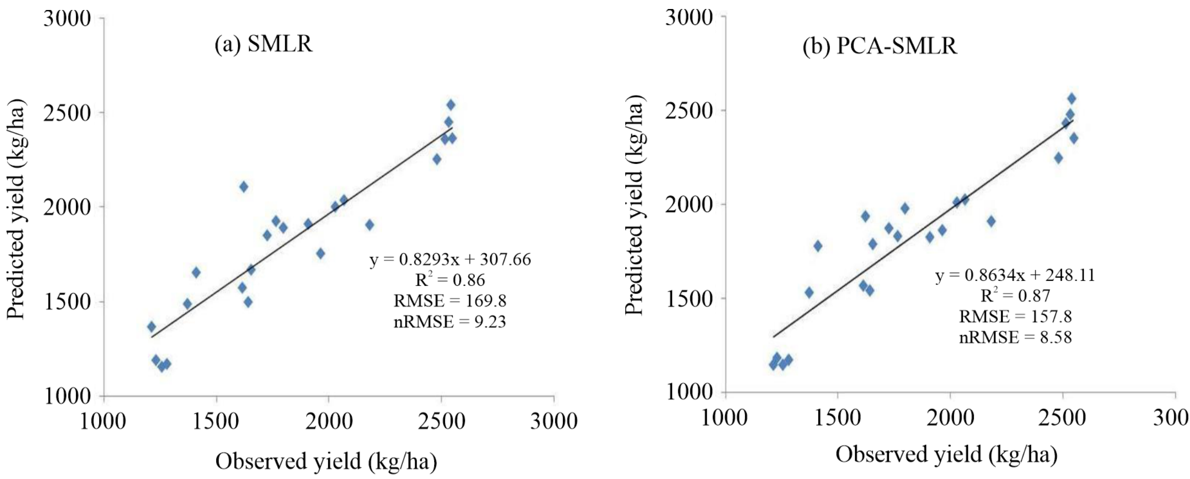


Fig. 2. Performance during calibration of model developed using (a) SMLR and (b) PCA-SMLR techniques for mustard yield prediction at vegetative stage

Performance of the model during calibration developed for mustard yield prediction at vegetative stage using SMLR and PCA-SMLR techniques are shown in Fig. 2. The RMSE value during calibration was lower for PCA-SMLR modal (157.8 kg/ha) followed by SMLR (169.8 kg/ha). During calibration nRMSE value was < 10% for both the models having lower value 8.58% for PCA-SMLR followed by 9.23% for SMLR model. The coefficient of determination (R^2) was significant at 1% probability level for all developed models. Value of coefficient of determination R^2 for models developed by different techniques for estimating the mustard crop yield at vegetative stage was 0.86% for model developed by SMLR techniques and 0.87% for modal developed by PCA-SMLR techniques. Performance of modal developed for mustard yield prediction at vegetative stage using SMLR and PCA-SMLR techniques during validation are shown in Fig. 3. During validation nRMSE value was lower 12.33% for PCA-SMLR followed by 12.38% for SMLR model. The RMSE value during validation was lower for PCA-SMLR modal (323.0 kg/ha) followed by SMLR (324.5 kg/ha). The most important weather parameter identified by SMLR for mustard yield prediction at vegetative stage is time and Z21 (weighted minimum temperature). Equations developed for mustard crop yield prediction at vegetative stage by SMLR is given in Table 1.

Percentage deviation of predicted yield for mustard crop done at vegetative stage by observed

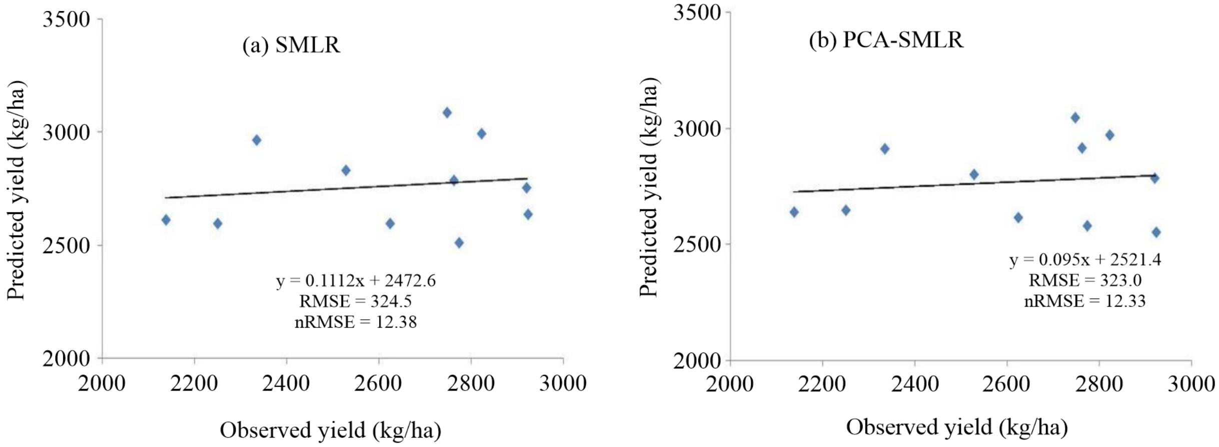


Fig. 3. Performance during validation of model developed using (a) SMLR and (b) PCA-SMLR techniques for mustard yield prediction at vegetative stage

Table 1. Mustard Yield prediction at different growth stage during *Rabi* 2018-19 and 2019-20

Model	Model equation	Predicted yield (kg/ha)		Observed Yield (kg/ha)		Percentage deviation	
		2018-19	2019-20	2018-19	2019-20	2018-19	2019-20
At Vegetative stage							
SMLR	y=1854.46+61.14*time+Z21*91.35	3088.8	3266.5	2725.6	2746.6	13.33	18.93
PCA-SMLR	y=1176.57+time*54.54+PC5*126.82	2898.8	2751.4	2725.6	2746.6	6.35	0.17
At flowering stage							
SMLR	y=1933.2+57.17*time+83.8*Z21	3061.7	3127.6	2725.6	2746.6	12.33	13.87
PCA-SMLR	y=1185.674+time*53.84+PC5*129.36	2984.7	2920.1	2725.6	2746.6	9.51	6.32
At grain filling stage							
SMLR	y=1656.66+57.15*time+78.95*Z21	2996.5	3155.4	2725.6	2746.6	9.94	14.88
PCA-SMLR	y=1168.9+time*55.12+PC5*142.85	2898.0	2962.9	2725.6	2746.6	6.33	7.88

yield for ICAR-IARI, New Delhi during *Rabi* 2018-19 and 2019-20 are shown in Table 1. During Rabi 2019-20 percentage deviation of predicted yield by observed yield was 0.17% for PCA-SMLR and 18.93% for SMLR model respectively. During *Rabi* 2018-19 the percentage deviation was lower 6.35% for PCA-SMLR and 13.33% for SMLR modal respectively. Kumar *et al*.(1999) developed stepwise regression technique to predict pigeon pea yield in Varanasi district using different weather variables, appropriate weighted and un-weighted weather indices. SMLR used for pre-harvest wheat crop yield estimation because of its more consistent performance and applicability at zone or state level (Garde *et al*., 2015). Feature selection helps to attain selection of best regression variables and thereby good interpretable results among independent variables (Singh *et al*., 2014).

Mustard Yield prediction by SMLR and PCA-SMLR model at flowering stage

Model for Mustard yield prediction at flowering stage for ICAR-IARI, New Delhi was developed by SMLR and PCA-SMLR using long term crop yield data as well as long period daily weather data during sowing to flowering stage (from 40th to 4th standard meteorological week).

Model performance during calibration for mustard yield prediction at flowering stage developed by SMLR and PCA-SMLR techniques are shown in Fig. 4. The RMSE value during calibration was lower for PCA-SMLR modal (161.2 kg/ha) followed by SMLR (164.6 kg/ha). Value of coefficient of determination R^2 for models developed by different techniques for estimating the mustard crop yield at flowering stage was 0.87% for model

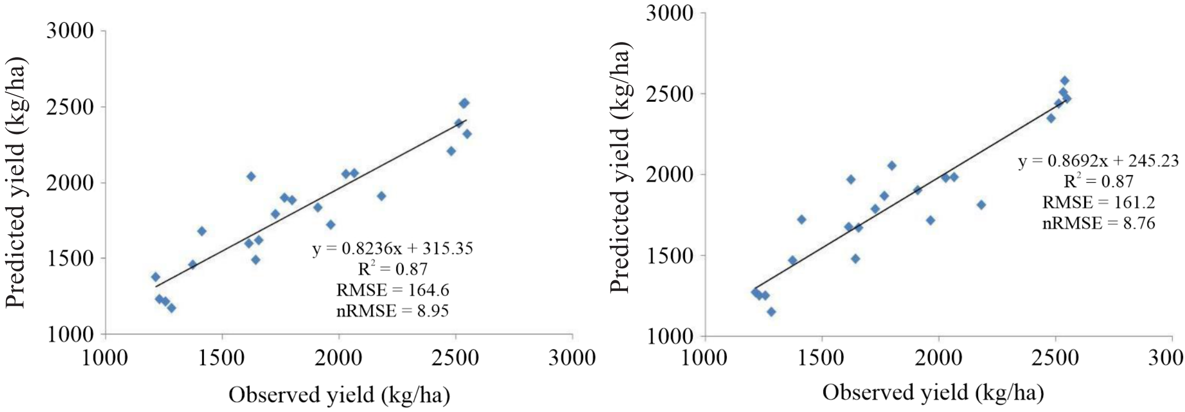


Fig. 4. Performance during calibration of model developed using (a) SMLR and (b) PCA-SMLR techniques for mustard yield prediction at flowering stage

developed by SMLR and PCA-SMLR techniques. During calibration nRMSE value was < 10% for both the models having lower value 8.76% for PCA-SMLR followed by 8.95% for SMLR. Performance of model developed for mustard yield prediction at flowering stage using SMLR and PCA-SMLR techniques during validation are shown in Fig. 5. During validation nRMSE value was lower 11.29% for PCA-SMLR followed by 11.55% for SMLR model. The RMSE value during validation was lower for PCA-SMLR modal (295.7 kg/ha) followed by SMLR (302.7 kg/ha). The most important weather parameters identified by SMLR for mustard yield prediction at flowering stage are time and Z21 (weighted minimum temperature). Equation developed for mustard crop yield prediction at flowering stage by SMLR is given in Table 1.

Percentage deviation of predicted yield for mustard crop done at flowering stage by observed yield for ICAR-IARI, New Delhi during *Rabi* 2018-19 and 2019-20 are shown in Table 1. During *Rabi* 2019-20 percentage deviation of predicted yield by observed yield was 6.32% for PCA-SMLR and 13.87% for SMLR model. During *Rabi* 2018-19 the percentage deviation was lower 9.51% for PCA-SMLR model followed by 12.33% for SMLR modal respectively. Lower value of percentage deviation was observed by PCA-SMLR as compared to corresponding value by SMLR in both the year. Vashisth *et al*. (2018) reported that percentage deviation of estimated yield by actual yield of maize crop done at flowering stage and at grain filling stage was 10.3 and 7.1% by weather based statistical model.

Mustard Yield prediction by SMLR and PCA-SMLR model at grain filling stage

Model for Mustard yield prediction at grain filling stage for ICAR-IARI, New Delhi was developed by SMLR and PCA-SMLR using long term crop yield data as well as long period daily weather data during sowing to vegetative stage (from 40th to 8th standard meteorological week).

Model performance during calibration for mustard yield prediction at grain filling stage developed by SMLR and PCA-SMLR techniques are shown in Fig. 6. The RMSE value during calibration was 155.5 kg/ha for PCA-SMLR and 159.7 kg/ha for SMLR. During calibration nRMSE value was <10% for both the models having lower value 8.45% for PCA-SMLR and 8.68% for SMLR model. Value of coefficient of determination R^2 for models developed for estimating the mustard crop yield at grain filling stage was 0.87% for SMLR model and 0.88% for PCA-SMLR modal.

Performance of the model developed for mustard yield prediction at grain filling stage using SMLR and PCA-SMLR techniques during validation are shown in Fig. 7. During validation nRMSE value was lower 11.43% for PCA-SMLR followed by 12.13% for SMLR model. The RMSE value during validation was lower for PCA-SMLR modal (299.4 kg/ha) followed by SMLR (317.8 kg/ha). The most important weather parameters identified by SMLR for mustard yield prediction at grain filling stage are time and Z21 (weighted minimum temperature). Equation developed for mustard crop yield

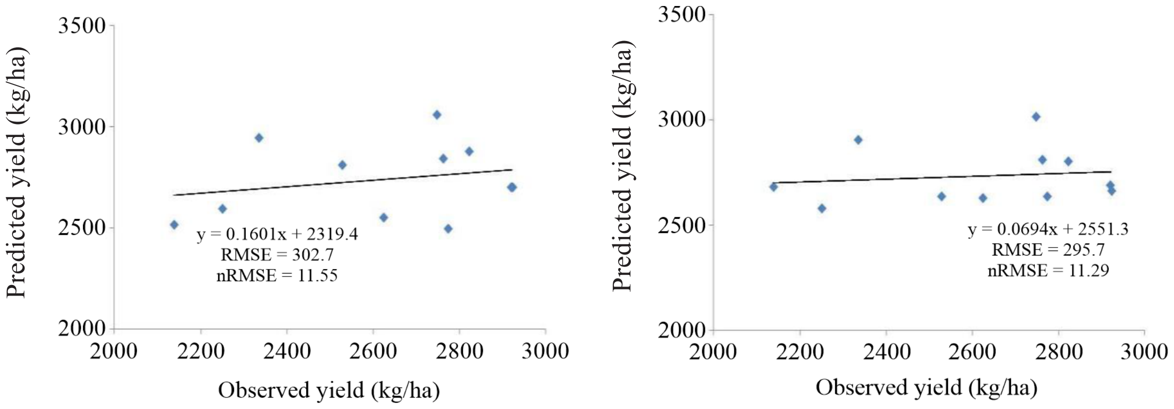


Fig. 5. Performance during validation of model developed using (a) SMLR and (b) PCA-SMLR techniques for mustard yield prediction at flowering stage

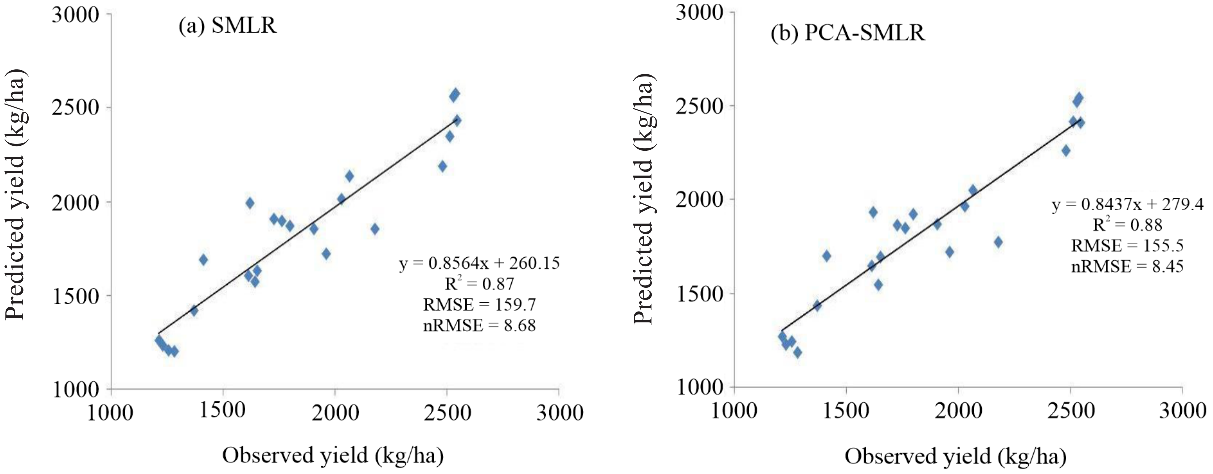


Fig. 6. Performance during calibration of model developed using (a) SMLR and (b) PCA-SMLR techniques for mustard yield prediction at grain filling stage

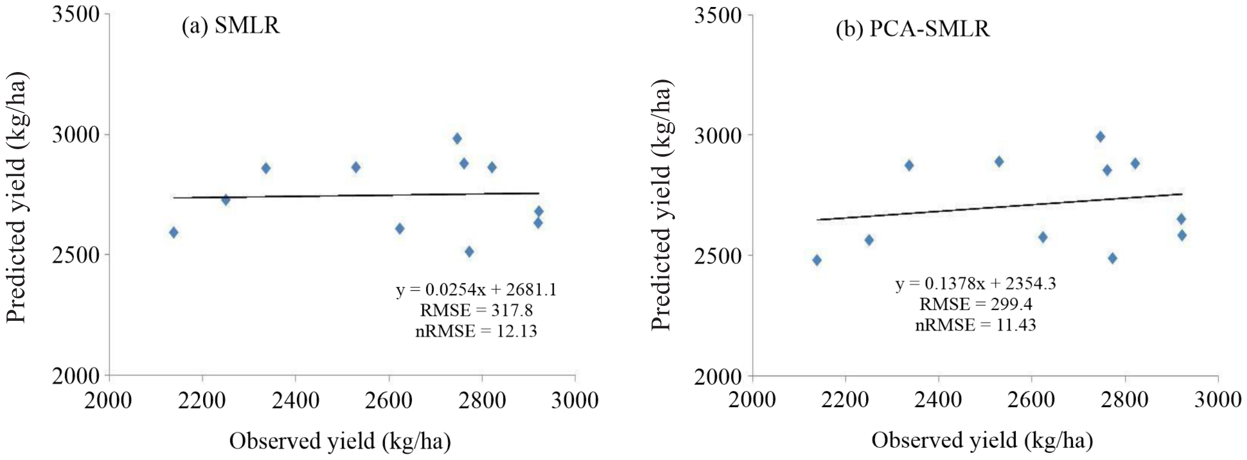


Fig. 7. Performance during validation of model developed using (a) SMLR and (b) PCA-SMLR techniques for mustard yield prediction at grain filling stage

prediction at grain filling stage by SMLR is given in Table 1.

Percentage deviation of predicted yield for mustard crop done at grain filling stage by observed yield for ICAR-IARI, New Delhi during *Rabi* 2018-19 and 2019-20 are shown in Table 1. During *Rabi* 2019-20 percentage deviation of predicted yield by observed yield was 7.88% for PCA-SMLR and 14.88% for SMLR model respectively. During *Rabi* 2018-19 the percentage deviation was lower 6.33% for PCA-SMLR model followed by 9.94% for SMLR modal. Lower value of percentage deviation was observed by PCA-SMLR as compared to corresponding value by SMLR in both the year. Vashisth *et al*. (2014) reported that percentage deviation of observed yield by estimated yield done at forty-five days before harvest by weather based statistical model was found to be 10.7, 5.7 and 8.53 respectively during the period of 2011-12, 2012-13 and 2013-14. Similarly, the percentage deviation of yield prediction done at 25 days before harvest by weather based statistical model was 9.7, 7.0 and 8.29 respectively. Singh *et al*. (2014) reported that statistical models based on weather indices can successfully simulate multi-stage yield forecast of wheat at mid-season and at pre-harvest for Amritsar, Bhatinda and Ludhiana districts. This model is simple, does not require any sophisticated statistical tools, and can be used satisfactorily for district, agro-climatic zone and state level forecasting.

Conclusions

Based on the overall performance of models developed for mustard yield prediction at different growth stage PCA-SMLR model performed better than SMLR model having lower value of nRMSE and RMSE for both the year for both the year. Percentage deviation of observed yield by predicted yield done at different growth stage using PCA-SMLR models had lower value as compared to corresponding value by SMLR model for both the year. From this study it may be concluded that PCA-SMLR and SMLR model based on weather parameters can be used for district level yield prediction at different crop growth stage and PCA-SMLR performing better than SMLR model.

References

Aravind, K.S., Vashisth, Ananta, Krishanan, P. and Das, B. 2022. Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *Journal of Agrometeorology* **24**(1): 18-25.

Azfar, M., Sisodia, B.V.S., Rai, V.N. and Devi, M. 2015. Pre-harvest forecast models for rapeseed & Maize yield using principal component. *Mausam* **4**: 761-766.

Garde, Y.A., Dhekale, B.S. and Singh, S. 2015. Different approaches on pre harvest forecasting of wheat yield. *Journal of Applied and Natural Science* **7**(2): 839-843.

Kumar, R., Gupta, B.R.D., Athiyaman, B., Singh, K.K. and Shukla, R.K. 1999. Stepwise regression technique to predict pigeon pea yield in Varanasi district. *Journal of Agrometeorology* **1**(2): 183-186.

Singh, R.S., Patel, C., Yadav, M.K. and Singh, K.K. 2014. Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *Journal of Agrometeorology* **16**: 199-202.

Singh, A.K., Vashisth, Ananta, Sehgal, V.K, Goyal, A., Pathak, H. and Parihar, S.S. 2014. Development of Multi Stage District Level Wheat Yield Forecast Models. *Journal of Agricultural Physics* **14**(2): 189-193.

Vashisth, Ananta, Singh, R. and Choudhry, Manu. 2014. Crop yield forecast at different growth stage of wheat crop using statistical model under semi arid region. *Journal of Agro ecology and Natural Resource Management*: 1-3.

Vashisth, Ananta, Goyal, A. and Roy, Debasish. 2018. Pre harvest maize crop yield forecast at different growth stage using different model under semi arid region of India. *International Journal of Tropical Agriculture* **36**(4): 915-920.

Vashisth, Ananta and Aravind, K.S. 2020. Multistage Mustard Yield Estimation Based on Weather Variables using Multiple Linear, LASSO and Elastic Net Models for Semi Arid Region of India. *Journal of Agricultural Physics* **20**(2): 213-223.

Verma, U., Piepho, H.P. and Goyal, A. 2016. Role of climatic variables and crop condition term for mustard yield prediction in Haryana. *International Journal of Agricultural and Statistical Science* **12**: 45-51.

Received: 2 September 2023; Accepted: 18 December 2023